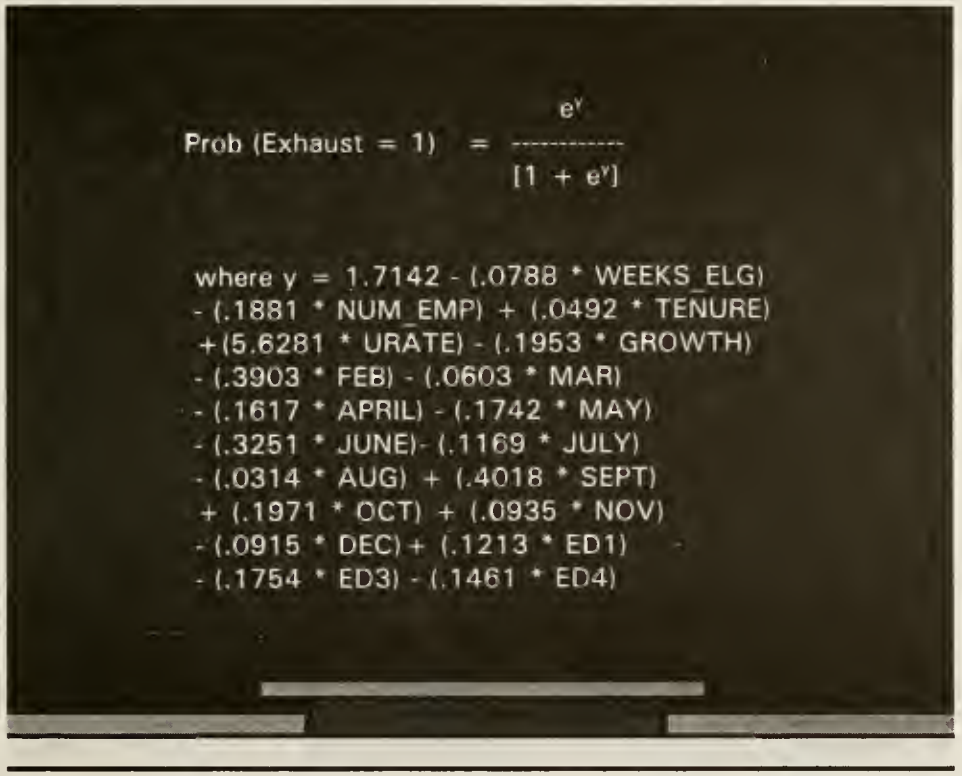
S
368.44
L1smui
1996
Montana. Dept. of
Labor and Industry
A statistical
model for UI
profiling in
Montana

# A STATISTICAL MODEL FOR UI PROFILING IN MONTANA


Prob (Exhaust = 1) = e^y / [1 + e^y]
where y = 1.7142 - (.0788 * WEEKS_ELG)
- (.1881 * NUM_EMP) + (.0492 * TENURE)
+ (5.6281 * URATE) - (.1953 * GROWTH)
- (.3903 * FEB) - (.0603 * MAR)
- (.1617 * APRIL) - (.1742 * MAY)
- (.3251 * JUNE) - (.1169 * JULY)
- (.0314 * AUG) + (.4018 * SEPT)
+ (.1971 * OCT) + (.0935 * NOV)
- (.0915 * DEC) + (.1213 * ED1)
- (.1754 * ED3) - (.1461 * ED4)

STATE DOCUMENTS COLLECTION

JUL 03 1996

MONTANA STATE LIBRARY
1515 E. 6th AVE.
HELENA, MONTANA 59620

**Published April 1996**
**by the**
**Montana Department of Labor & Industry**
**P. O. Box 1728**
**Helena, Montana 59624**
**Phone: (406) 444-2430**

PLEASE RETURN

Montana State Library
3 0864 1006 0229 4

# A STATISTICAL MODEL FOR UI PROFILING IN MONTANA

This report was developed as part of the partnership agreement for worker profiling and reemployment services between the U.S. Department of Labor, Employment and Training Administration and the Montana Department of Labor & Industry (Grant # M-4974-5-00-97-75). The Department's Research and Analysis Bureau was given the responsibility of developing this statistical model. Bob Liffring, state statistician, and Phil Brooks, state economist, were the principal authors.

**Published April 1996**
**by the**
**Montana Department of Labor & Industry**
**P. O. Box 1728**
**Helena, Montana 59624**
**Phone: (406) 444-2430**

# TABLE OF CONTENTS


| | |
|---|---|
| Executive Summary | 1 |
| Main Report | 2 |
| Introduction | 2 |
| Overall Approach | 2 |
| The Model | 3 |
| Validation of the Model | 5 |
| Sensitivity Analysis | 6 |
| Appendix A--Variables considered for Montana UI Profiling | 7 |
| Variables found in the Model (Equation 1-1) | 7 |
| Other Variables--Those not found in the Model (Equation 1-1) | 16 |
| Appendix B--Univariate Analysis of Selected Categorical Variables | 26 |
| Appendix C--The Multivariate Model | 29 |
| Preliminary Analyses | 29 |
| The Final Model | 29 |
| Investigation of Interaction Terms | 30 |
| Validation of the Model | 31 |
| Appendix D--Sensitivity Analysis of the Recommended UI Profiling Model to Changes in the Values of the Predictors | 33 |
| References | 35 |

# EXECUTIVE SUMMARY

Public Law 103-152 requires that each state develop and implement a system to identify those unemployment insurance claimants who will probably exhaust their benefits and need extra job search assistance. Part of this effort involves specifying a statistical model, or mathematical equation, which assigns a probability (chance) of exhaustion to each individual. By using this equation, the probability of new claimants exhausting their benefits will be calculated using their individual characteristics. Those with the highest probability of exhaustion will be given services beyond those given to all claimants.

In order to develop this model, a number of factors that could possibly influence the chance that a particular claimant would exhaust his/her benefits were formulated and tested. The unemployment insurance claimant profiling work completed for the nation and other states was important in constructing this list of potential explanatory factors. Also helpful were analyst knowledge of the workings of the Montana economy, and information from program managers concerning the unemployment insurance system itself.

Thirty-four different factors were tested for their strength in explaining exhaustion of benefits. These factors included claimant's personal characteristics (education, occupation, etc.); claimant's employment history (base wages, weekly unemployment benefit, years employed with the last employer, etc.); employer traits (growth or nongrowth industry, seasonal industry, size of business, etc.); and economic characteristics of the claimant's resident county (unemployment rate, per capita income, etc.). Three potentially useful explanatory factors (age, race, sex) were not available for use in the model because of legal mandates.

The factors which proved to be the best together in explaining claimants exhausting their benefits, were a small group from the above described list. These were:

1. The number of weeks a claimant is eligible for benefits;

2. The number of companies employing the claimant during the base period;

3. The number of years a claimant was employed by his/her last employer;

4. The previous year's unemployment rate in the claimant's resident county;

5. Whether or not the claimant's last employer was part of a growth industry;

6. The claimant's level of education;

7. The month in which the initial claim was filed, a seasonal factor.

Using this model, the identification of claimants that are expected to exhaust their benefits will improve by about 22 percent over random selection methods.

# A STATISTICAL MODEL FOR UI PROFILING IN MONTANA

## INTRODUCTION

Public Law 103-152 requires each state develop and implement a system to identify those unemployment insurance claimants who will probably exhaust their benefits and need extra job search assistance. Part of this effort involves specifying a statistical model, or mathematical equation, which assigns a probability of exhaustion to each individual. This equation is to be developed using a procedure called "logistic regression."

By using this equation, the probability of each new claimant exhausting his or her benefits will be calculated using his or her individual characteristics. Those with the highest probability of exhaustion will be given services beyond those given to all claimants.

The purpose of this paper is to present Montana's model, to document how it was developed, and to describe some of the factors which seem to influence the exhaustion of benefits.

## OVERALL APPROACH

Data used in developing this model comes from three databases. They are:

- o The Benefit Automation Rewrite (BEAR) Information System which serves as the primary source of data. BEAR provides basic information on the claimant, benefits received, his/her employment history, and his/her employers.

- o The ES 202 Program (Employment, Wages and Contributions Report) which supplies additional, more detailed, information about the claimant's employers.

- o The Labor Market Information (LMI) program which generates geographical and economic information. Two primary data sets are used in this effort. They are annual unemployment statistics from the Local Area Unemployment Statistics (LAUS) program and industry growth projections.

The Montana Job Service database has also been examined for possible information and possible predictors, but was found to be able to contribute nothing more to the model.

The variables examined in this effort are defined and described in Appendix A. An analysis of each, when used alone, to predict exhaustion is presented in Appendix B. Such analyses constitutes the first step in statistical modelling and serves as a method to identify the most promising predictors. It also provides insights into the factors which influence exhaustion.

The overall approach to developing this equation is outlined in Table 1. Data from 1994 UI claims was used to develop the equation. Data from 1993 and 1992, respectively, validated the model. The

number of records in each database, the proportion of exhaustees, and the number of usable records are shown in Table 1.

**Table 1**
**General Strategy used in Developing**
**a Statistical Model to Support UI Profiling in Montana**

| YEAR | DATABASE USED | NO. RECORDS DOWNLOADED | NO. WITH OUT OF STATE RESIDENCE | NO. WITH OUT OF STATE EMPLOYERS | NO. RECORDS USABLE | % NOT USABLE | PROBABILITY OF EXHAUSTION |
|---|---|---|---|---|---|---|---|
| 1994 | Model Development | 9,112 | 11 | 314 | 8,787 | 3.6% | 48.0% |
| 1993 | Model Validation | 9,360 | 19 | 50 | 9,291 | 0.7% | 51.3% |
| 1992 | Model Validation | 10,178 | 29 | 824 | 9,325 | 8.4% | 52.2% |

All records from these three years were screened to ensure they satisfied the following characteristics:

- o All records were for claimants who were awarded UI benefits
- o All of these claimants were not a member of a Union with a hiring hall, and
- o All were not scheduled to be recalled (i.e. returned to work) at some fixed future time.

The UI program will continue to screen for these characteristics. Because the logistic regression model requires that all data elements be present to work, it is recommended that two additional screens be added. They are:

- o All claimants to be profiled must have a Montana residence. (State FIPS Code = 30). Montana Job Service will not deliver services to people no longer living in the state.
- o The claimant's last employer must have a Montana UI Account Number. That is, they must be a Montana employer. Not having the employer's UI Account Number means that all predictor variables based on employer and industry are missing.

## THE MODEL

Table 2 contains the equation (Equation 1-1) being recommended for UI profiling in Montana. The y term, made up of known values of the predictor variables is called the "logit." By putting the characteristics of a single individual into equation 1-1, one obtains a probability, a value between 0 and 1. The higher the probability, the greater the chance that this particular person will exhaust benefits. Consequently, the application of this model to the characteristics of several individuals can be used to prioritize their relative risk of exhaustion.

## Table 2
## The Proposed Statistical Model

$$\text{Prob (EXHAUST = 1)} = \frac{e^{y}}{[1 + e^{y}]} \qquad \text{(EQUATION 1-1)}$$

where y = 1.7142 - (.0788 * WEEKS_ELG) - (.1881 * NUM_EMP)
+ (.0492 * TENURE) + (5.6281 * URATE) - (.1953 * GROWTH)

- (.3903 * FEB) - (.0603 * MAR) - (.1617 * APRIL) - (.1742 * MAY)
- (.3251 * JUNE) - (.1169 * JULY) - (.0314 * AUG) + (.4018 * SEPT)
+ (.1971 * OCT) + (.0935 * NOV) - (.0915 * DEC)

+ (.1213 * ED1) - (.1754 * ED3) - (.1461 * ED4)

and where

**EXHAUST** is the act of the claimant having used up all UI eligible benefits in the benefit year. That is, benefits are exhausted when the amount paid is equal to the eligible amount. **EXHAUST** takes on a value of 1 if the claimant has exhausted benefits, and a value of 0 if not.

**WEEKS_ELG** is the number of weeks for which the claimant is eligible for Unemployment Insurance benefits based on wages earned during his/her base period. This ranges from 8 to 26 and is always an even number.

**NUM_EMP** is the number of employers the claimant has had during his/her base period of UI eligibility. In theory, this number may range from 1 to infinity, but in practice numbers greater than 5 are uncommon.

**TENURE** is the number of years the claimant has been employed by his/her last employer. It is calculated by subtracting the beginning date of employment from the last day of employment and dividing by 365.

**URATE** is the previous year's unemployment rate in the claimant's county of residence.

**GROWTH** is equal to 1 if the industry in which the claimant last worked (based on his/her last employer's SIC Classification) is projected to generate 50 or more jobs in Montana annually in the next 5 years, and equal to 0 otherwise.

**FEB, MAR, APRIL, MAY, JUNE, JULY, AUG, SEPT, OCT, NOV & DEC** indicate the month in which the initial UI claim was filed. All are dummy variables, taking on the value of one if the month is a hit and zero otherwise. January (**JAN**) is used as the reference for the other months.

**ED1, ED3, & ED4** are dummy variables to signify the highest educational level achieved by the claimant. All three variables take on values of either zero or one. **ED1** indicates no high school diploma; **ED3**, some college; and **ED4**, a college degree. Claimants with a high school education (**ED2**) are used as the baseline for the other three variables.

The model uses 19 predictors or independent variables. Three of these are categorical variables which describe the claimant's formal educational attainment; eleven others depict the month in which the claim was filed. The monthly variables are used to isolate some of the factors associated with seasonal jobs. Not all of the educational and monthly variables are in themselves statistically significant. However, in logistic regression, the use of such categorical variables requires that if one is used, they all must be used.

The technical aspects of the model development are presented in Appendix C. This appendix also contains a list of the statistical properties of each of the variables contained in the final model, the results of a limited investigation of some of the interactions between selected predictors, and a discussion of the validation process.

## VALIDATION OF THE MODEL

A Chi-squared test indicates that, overall, Equation 1-1 fits the 1994 data reasonably well. The theory behind this model is based upon correctly describing the outcome variable, exhaustion. This test indicated this particular configuration of data would appear randomly only 8% of the time, implying that the model is a reasonable fit for the data.

On the other hand, UI profiling is more interested in predicting the exhaustees than the non-exhaustees. Consequently, one intuitively can validate the model by examining the proportion of exhaustees which are correctly identified. These percentages are compared to a baseline where no model is used.

Table 3 shows the results of such an examination. If the Montana Job Service were to have used Equation 1-1 to classify UI claimants from 1994 as potential exhaustees, they would have correctly identified 60.1% of them. The net result would have been 25.3% improvement over random selection. This, however, is using the same data to validate the model as was used to develop it. As such, this could be the best results we could hope for. Using Equation 1-1 on independent data from 1993 and 1992 show a 20.4% and 23.2% improvement, respectively. This suggests that when implemented, the model offers a 21.5% improvement over random selection.

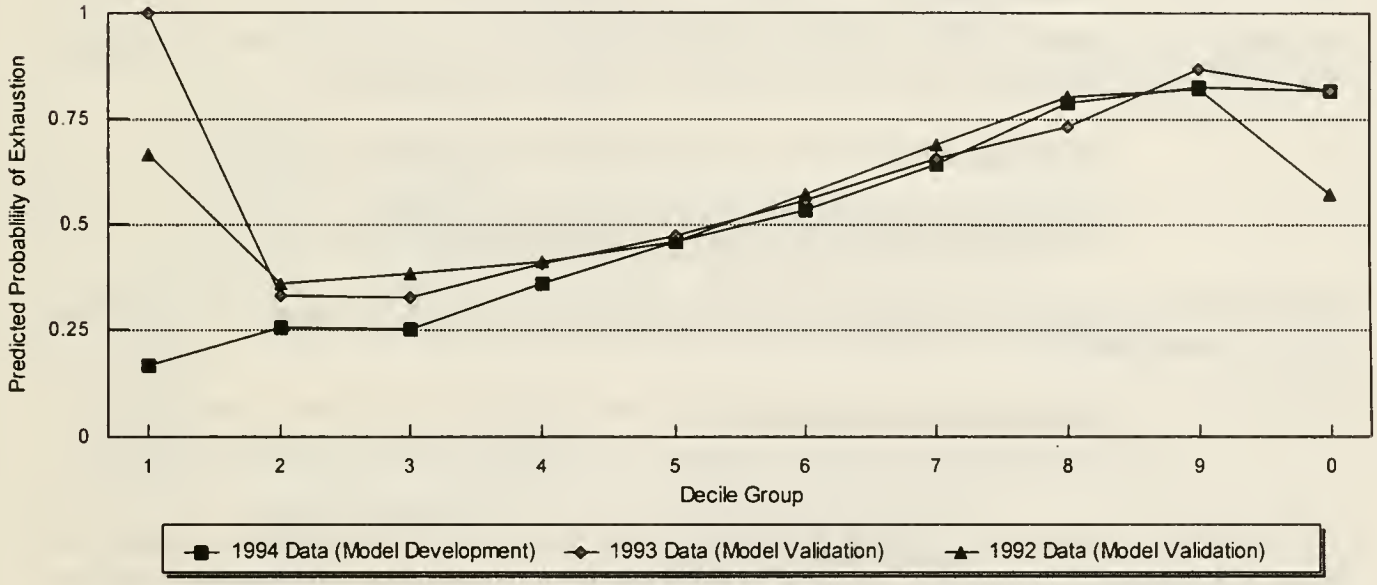
**Table 3**
**A Comparison of the Efficiency**
**of using the Statistical Model**
**to Predict Exhaustion in Different Years**

| | PROPORTION EXHAUSTEES | | % |
|---|---|---|---|
| YEAR | RANDOM SELECTION | PREDICTED | IMPROVEMENT |
| 1994 | 48.0% | 60.1% | 25.3% |
| 1993 | 51.3% | 61.8% | 20.4% |
| 1992 | 52.2% | 64.3% | 23.2% |

A more detailed examination of the predictive ability of the model is shown in Figure 1. This figure shows the range of predictions divided into deciles (tenths). A visual inspection of it shows that the model achieves a relatively good fit when the probability of exhaustion is between 20% and 80%, where most of the observations can be found. The fit is poor at the extremes, in the first and last decile ranges. However, very few of the observations are found here. In fact, there are only 11 cases for all three years which are in the first decile. The last decile contains 29 cases. Tables showing the actual data are found in Appendix C.

Of particular importance is the fact that the proportion of exhaustees monotonically increases by decile group.

**Figure 1**
**Probability of Exhaustion**
**by Decile Group**




## SENSITIVITY ANALYSIS

The change in the probability of exhaustion when changes were made in the predictor variables was examined in a limited fashion. Mean values of each of the predictors were used as a baseline. Results are presented in Appendix D.

# APPENDIX A
# VARIABLES CONSIDERED
# FOR
# MONTANA UI PROFILING

This appendix contains definitions and descriptions of the variables considered in developing the statistical model. It is separated into two parts: the first dealing with the variables found in the recommended model (Equation 1-1) and the second, with other variables considered and rejected. For obvious reasons the first part is more detailed.

Data from 1994 is used in all cases where the distribution of the data is displayed, discussed and described.

An analysis of the ability of each of these variables, expressed as categorical variables, to individually predict which claimants will exhaust benefits is contained in Appendix B.

## VARIABLES FOUND IN THE MODEL (EQUATION 1-1)

### EXHAUST—THE DEPENDENT (PREDICTED) VARIABLE

EXHAUST is a categorical variable which indicates if the claimant has exhausted his/her UI benefits. Its values are:

1 = Benefits have been exhausted.
0 = Benefits have not been exhausted.

The variable is constructed by comparing the maximum amount of benefits allowed (based upon payments made in the base period) with the total amount paid to the claimant. If the difference is 0, then the benefits have been exhausted.

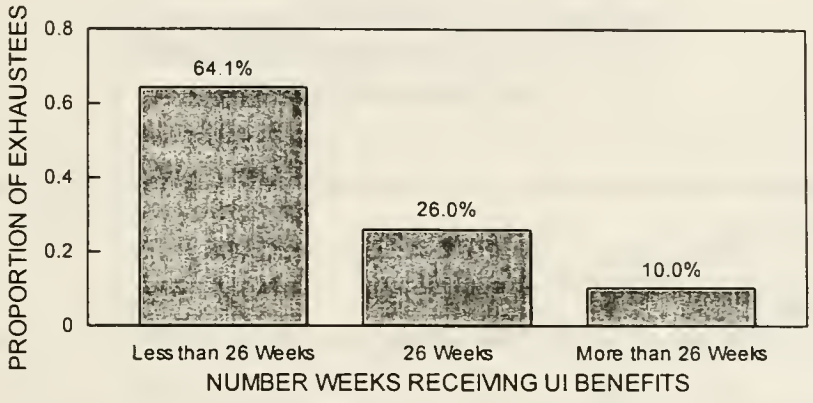
In the development of this statistical model, no attempt has been made to incorporate the amount of time it takes for a person to exhaust benefits. As can be seen in Figure A1, exhaustees can be separated into three separate groups based upon their weeks of eligibility:

- o Almost 65% of the Montana exhaustees used up their benefits before the allotted maximum length of 26 weeks. These were people whose income fluctuated widely from quarter to quarter in the base period.

- o About 25% of the people exhausted their benefits in precisely 26 weeks. These were people who had stable jobs and incomes during the base period.

- o Ten percent of the exhaustees lasted more than 26 weeks. These are primarily people who took part-time jobs and consequently didn't draw their maximum weekly amounts.

**Figure A1**
**Distribution of Claimants Exhausting Benefits**
**as a function of**
**the Number of Weeks they Received UI Benefits**




## THE INDEPENDENT (PREDICTOR) VARIABLES FOUND IN THE MODEL (EQUATION 1-1)

### THE NUMBER OF WEEKS OF UI ELIGIBILITY—WEEKS_ELG

The number of weeks of UI eligibility was not obtained directly from the database. Instead it was calculated using the following formula:

$$\text{Weeks Eligible} = \frac{\text{Total Eligible Benefits}}{\text{Weekly Benefit Amount}}$$

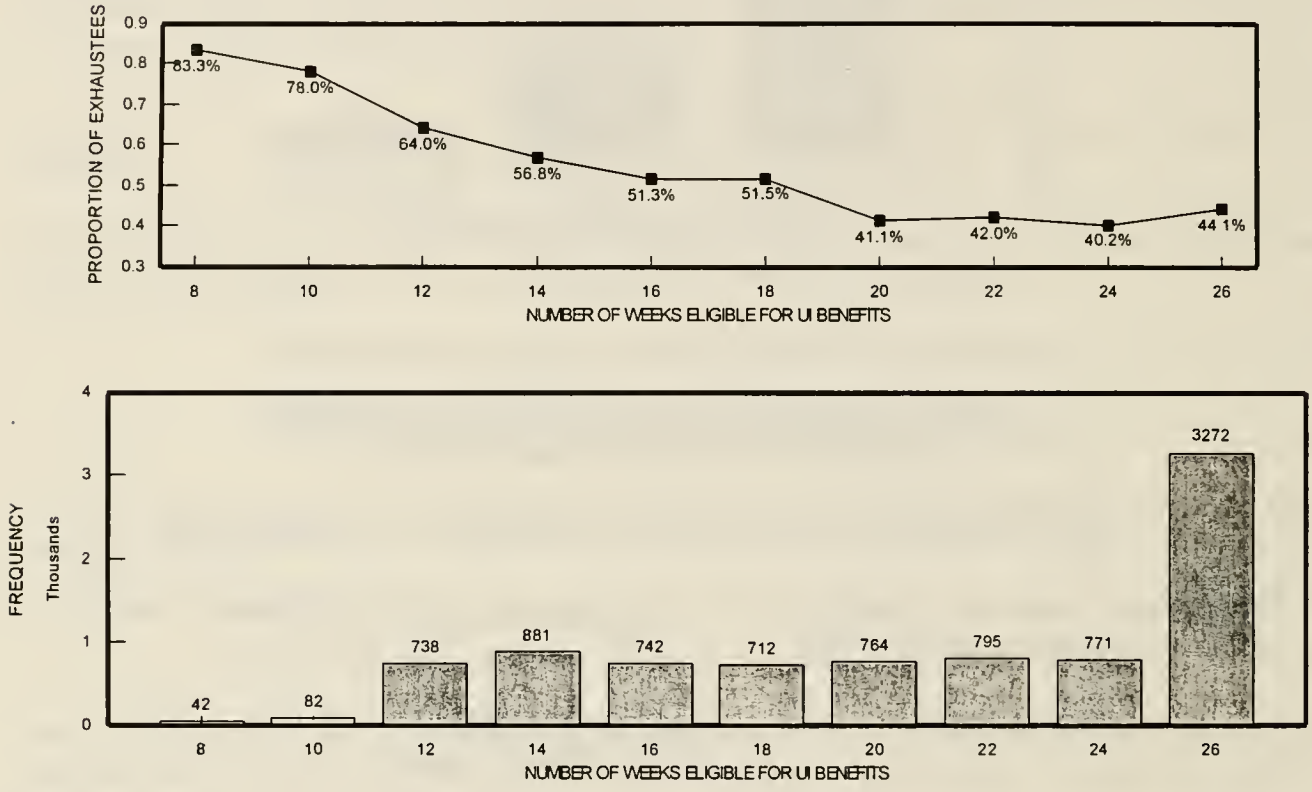
Since this statistic should be in the BEAR Database, it will be unnecessary to perform this calculation when the equation is used in profiling.

By definition, the number of weeks of eligibility has a limited range and is truncated at both ends. A claimant who qualifies for unemployment insurance is eligible for no less than 8 weeks and no more than 26. The weeks of eligibility increases in increments of 2 weeks. Because some claimants may work intermittently during their eligibility period, they may receive benefits for more than their allotted weeks of eligibility.

In the model, weeks of eligibility is treated as if it were a continuous variable. It is hypothesized that the number of weeks a claimant is eligible for benefits will be inversely related to the probability of exhaustion. Figure A2 shows this to be the case. Claimants with 8 weeks of eligibility experienced an 83.3% exhaustion rate. This rate of exhaustion slopes downward until one hits 20 weeks of eligibility. For 20 or more weeks of eligibility, that exhaustion rate is relatively static at 42%, ranging between 41.1% and 44.1%.

The bottom half of Figure A2 shows the distribution of claimants in terms of the number of weeks they are eligible for benefits. Almost one third of the claimants qualify for the full 26 weeks. It is evenly distributed between 12 and 24 weeks, with roughly 750 claimants qualifying for each two week increment.

**Figure A2**
**Proportion of Exhaustees as a Function of Weeks of Eligibility**



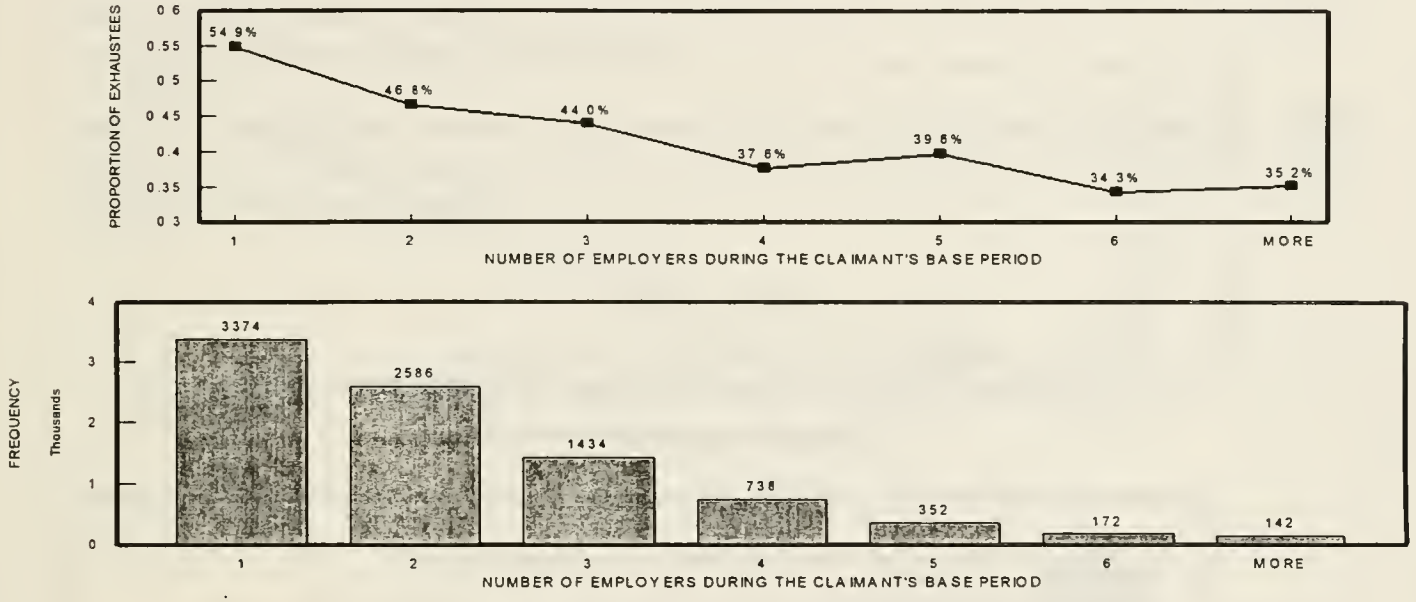

**NUMBER OF EMPLOYERS IN THE BASE PERIOD—NUM_EMP**

NUM_EMP is the number of base period employers who have paid UI contributions on behalf of the claimant. It is indicative of the total number of jobs held by the individual and of his/her attitude toward job hunting.

Although this is a whole number, NUM_EMP is treated in the model as if were continuous. It is argued that claimants with more than one base period employer are more aggressive in seeking employment and, consequently, less apt to exhaust their benefits. The top half of Figure A3 shows this to be the case, as the number of base employers increase, the probability of exhaustion decreases.

The bottom half of Figure A3 shows its distribution for exhaustees in 1994. One can see it is steadilydecreasing as the number of employers increase. Although not displayed, the distribution of non-exhaustees is similar in shape but much flatter.

**Figure A3**
**Proportion of Exhaustees as a Function of the Number of Base Period Employers**




## NUMBER OF YEARS EMPLOYED BY THE CLAIMANT'S LAST EMPLOYER—TENURE

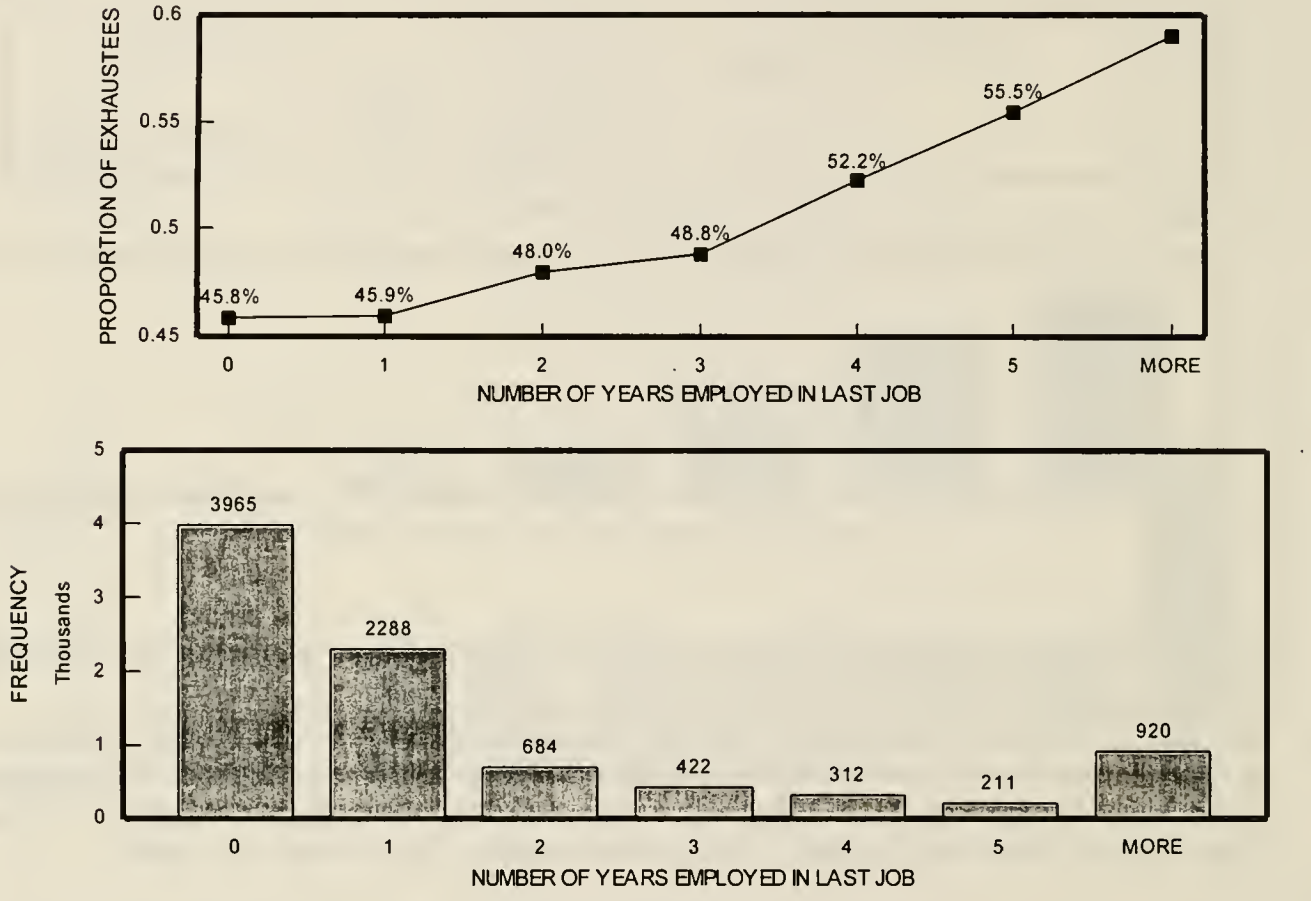
TENURE is the length of time that the claimant was employed by his/her last employer. Expressed in years, it is calculated by subtracting the start date of employment from the end date of employment and dividing by 365. Because of the way it is constructed, TENURE is a continuous variate. However, it is reported as if it were a discreet whole number with relatively few values.

Although the range of values for TENURE is greater than those for NUM_EMP, they are more concentrated. A person can work for a single company for 20, or even 30 years. However, the 1994 data shows that 45% of the claimants have worked less than a single year for the company laying them off. Ninety percent report that they have worked for their last company for five years or less. (See the bottom half of Figure A4.)

Other UI Profiling efforts have found that this variable, or some derivative of it has been "extremely important" in predicting an individual's chance of exhausting benefits. Reemployment difficulty increases with tenure. Theorists argue that TENURE measures job-specific human capital. Workers who have accumulated most of their qualifications while working for a single firm have developed skills that are uniquely valuable to that particular company. They have difficulty finding a demand for these same skills in other companies.

The 1994 UI claimant data seems to substantiate this reasoning, the top half of Figure A4 shows that the proportion of exhaustees increases as TENURE increases. Notice also that it is monotonically increasing.

**Figure A4**
**Proportion of Exhaustees as a Function of the Length of Time the Claimant was Employed by his/her Last Employer**




## WHETHER OR NOT THE CLAIMANT'S LAST EMPLOYER WAS IN A GROWTH INDUSTRY—GROWTH

GROWTH indicates if the claimant's last employer was considered to be in a growth industry. One would expect, people working in nongrowth, or declining, industries to have a tougher time finding a job after being laid off. That is, they would experience higher rates of exhaustion.

A growth industry in Montana has previously been defined as one which is expected to create at least 50 new jobs per year during the period between 1995 and 2000. This determination is based on a recent study which used 2-digit and 3-digit SIC codes to project industry growth.

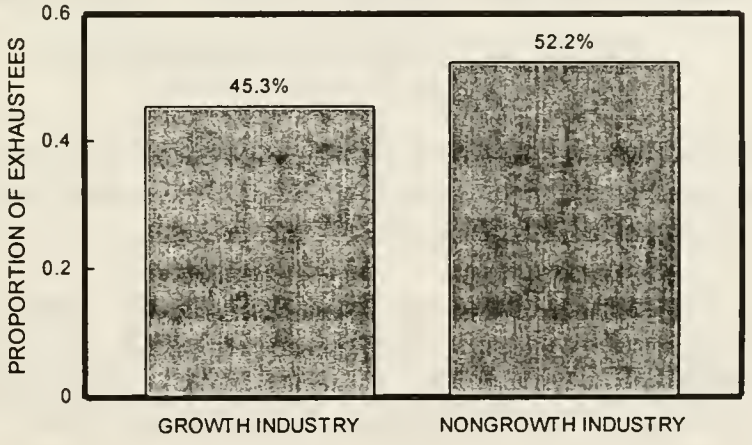
In the UI profiling project, this is a categorical variable with the following values:

1 = Claimant's last employer is in a growth industry in Montana.
0 = Claimant's last employer is in a declining or stagnant industry in Montana.

The expected increase in the number of jobs, expressed in percentage terms, could have been an alternative way of defining this variable. It is argued, however, that Montana contains many small industries, so percentages can be misleading. One might not consider an industry which employs 50 people and is expected to employ 75 within the next 5 years to be a growth industry. Yet, it would have a 50% growth rate for this same period.

Figure A5 compares the proportion of exhaustees found in growth and nongrowth industries in 1994. Although 45% of the exhaustees hail from growth industries, the chance of exhaustion is much higher in the nongrowth sector.

**Figure A5**
**Proportion of Exhaustees as a Function of whether or not the Claimant was Last Employed in a Growth Industry**



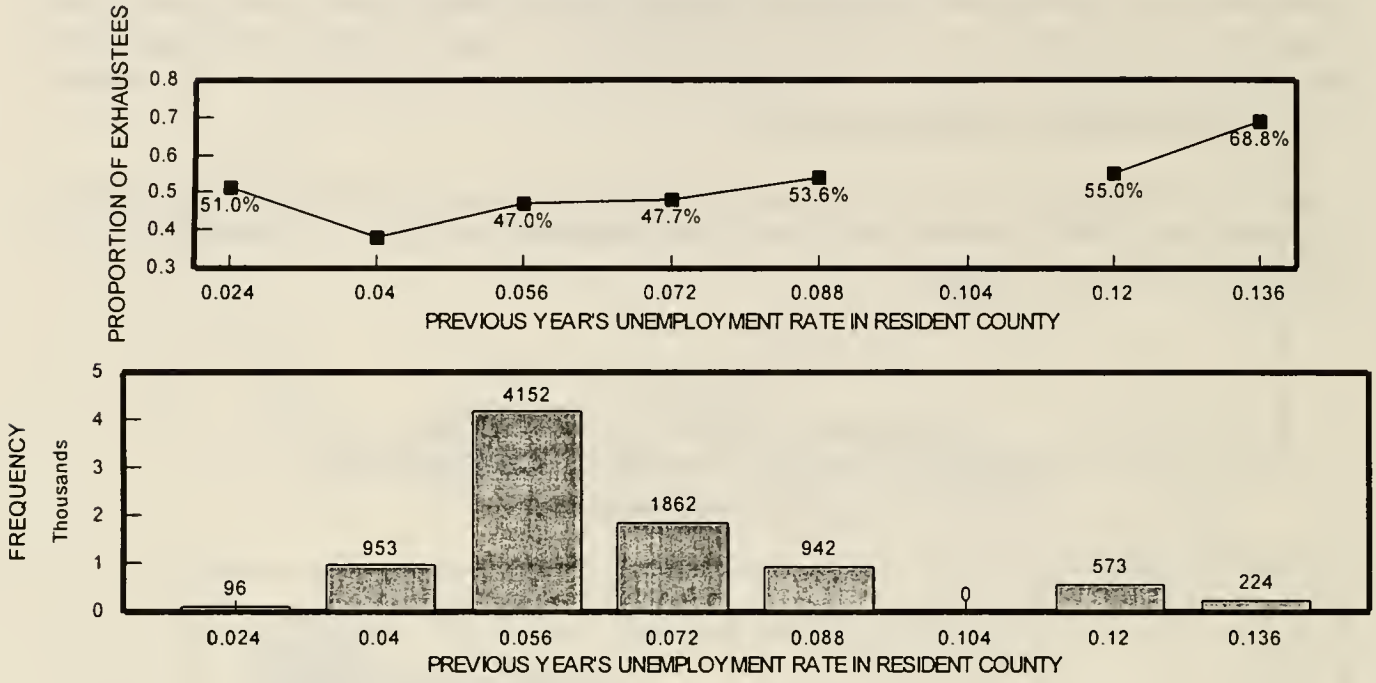

## UNEMPLOYMENT WITHIN THE CLAIMANT'S COUNTY OF RESIDENCE—URATE

URATE is defined as the previous year's unemployment rate for the claimant's county of residence. The county of residence is taken from the BEAR record; the unemployment rate is provided by the Local Area Unemployment Statistics (LAUS) program.

In theory, the proportion of exhaustees should increase as the local unemployment rate rises. That is, there will be higher rates of exhaustion in geographical areas where jobs are scarce. This phenomenon can be seen in Figure A6.

Other UI Profiling projects have shown that claimants who live in areas of high unemployment have a larger chance of exhausting their UI benefits than claimants who live in areas with lower unemployment rates.

**Figure A6**
**Proportion of Exhaustees as a Function of Local Unemployment Rates**




Because URATE is based on the county in which the claimant lives, and not on the individual, derived statistics are clustered around relatively few values. This takes away much of its predicting power. However, its impact is still significant.

An examination of 1994 records shows that URATE is roughly normally distributed (i. e. bell shaped) around the average of 6.4%. See the bottom of Figure A6. A problem arises, however, in that there are not observations for the range between 10% and 12%. The reason for this is that Montana only has 56 counties, thus URATE has only 56 possible values.

## EDUCATIONAL LEVEL

Although education is recorded in the BEAR database as the number of years of schooling completed by the claimant, it is used in this effort as a series of categorical variables. It is believed that the attainment of definitive goals, such as completing high school, is more important than the actual number of years spent in school. Consequently, the following categorical variables have been defined:

- **ED1**—Claimant has less than a high school education (no diploma). Corresponds to education values between 0 and 11. Roughly 10.0% of the 1994 UI claimants fell into this category.

o **ED2**—Claimant is a high school graduate. ED2 is comprised of those claimants with 12 years of schooling. They make up 57.4% of the population. ED2 is used as a base from which the other education variables are measured.

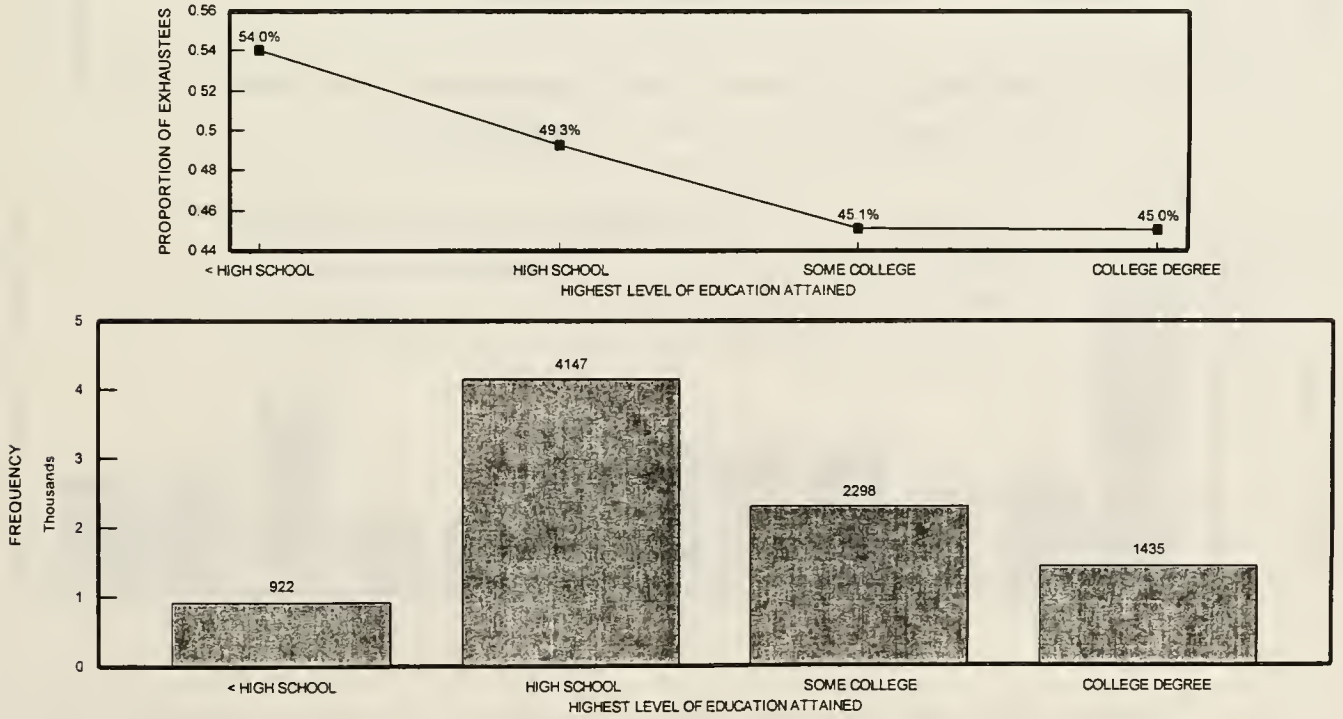
o **ED3**—Claimant has some college but did not graduate. This category corresponds to educational grade levels of 13, 14, & 15. Claimants in this category make up over a quarter of the population (26.2%).

o **ED4**—Claimant is a college graduate. It is assumed that people having completed 16 or more years of education are college graduates. No attempt has been made to distinguish between people holding a bachelor's degree from a master's or a PhD. College graduates make up approximately 16.4% of the 1994 population.

This variable, or some derivative of it, has been found in other UI profiling efforts to be highly sensitive to predicting an individual's chance of exhausting benefits. Researchers argue that persons with less education are more likely to exhaust their benefits. Other research, however, has found that both the less educated and the highly educated have more difficulty finding a job after being laid off. The reason for the latter is the highly educated may have become too specialized.

In Montana, exhaustion appears to be inversely related to educational attainment. As shown in the top half of Figure A7, the greatest percentage of exhaustees are found among the people who don't have a high school education. Contrary to findings in other states, college graduates do not have a greater difficulty in finding jobs after filing for unemployment insurance benefits.

**Figure A7**
**Proportion of Exhaustees as a Function of the Claimant's Educational Level**

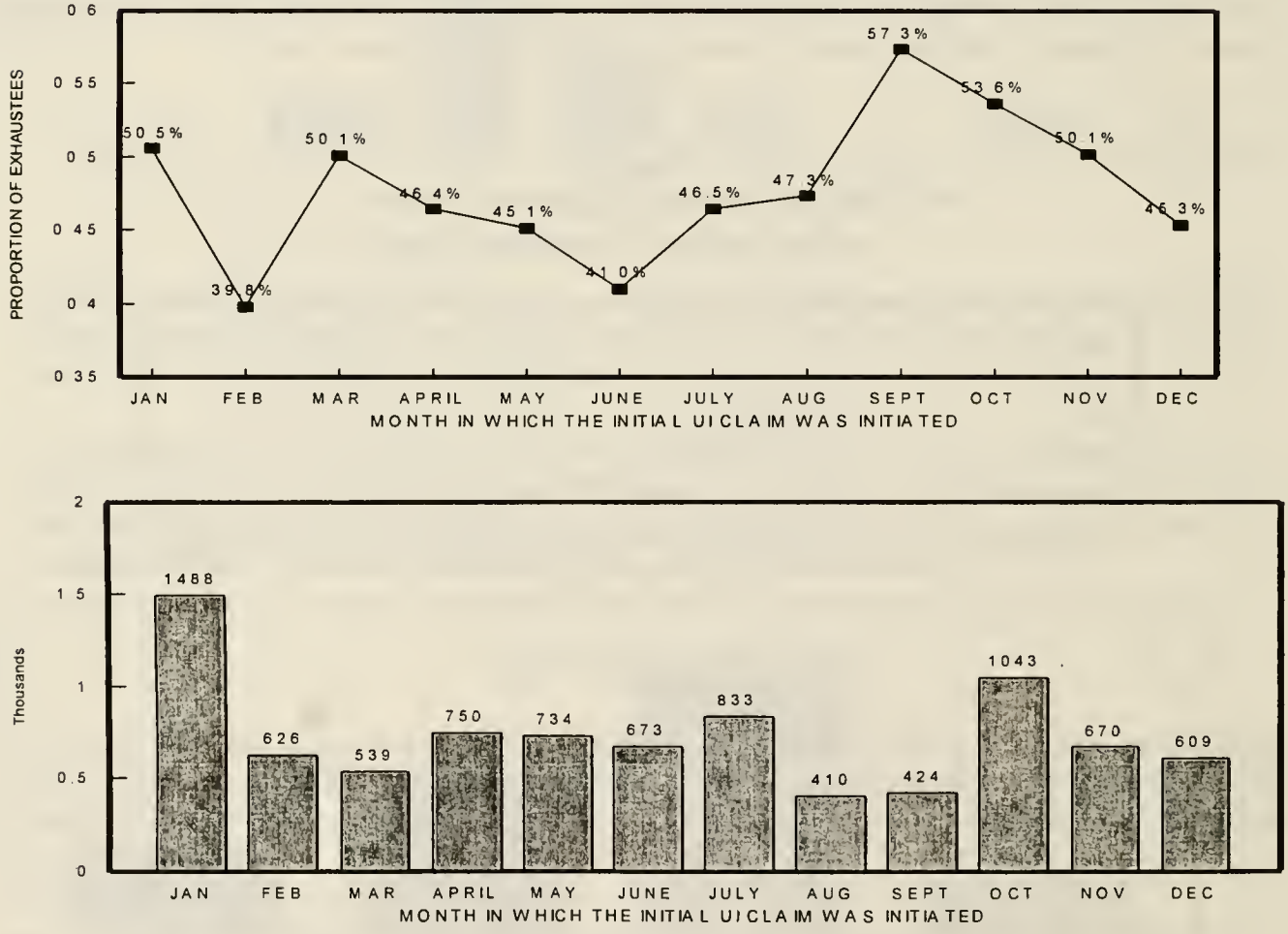


## SEASONALITY

Montana has many seasonal workers whose chosen lifestyle includes anticipated periods of unemployment during the year. These may be sugar beet processors who are generally employed during the sugar beet harvest (September through January), seasonal Forest Service and Bureau of Land Management personnel who are laid off during the winter months, school workers who have their summers off, and people associated with tourism who may be working during hunting, fishing and ski seasons. Statistically identifying such people is a difficult task.

The approach used in this project was to create 12 categorical variables which corresponded to the month in which the initial claim was filed. Each variable took the value of 1 if the claim was filed during the month being addressed, and zero otherwise. It is hypothesized that seasonal workers are less apt to exhaust their benefits.

**Figure A8**
**Proportion of Exhaustees as a Function of the Month in which the Claim was Filed**



## OTHER VARIABLES—THOSE NOT FOUND IN THE MODEL (EQUATION 1-1)

The analysis of the potential influence of these variables upon **EXHAUST** (the dependent variable) can be found in Appendix B. This section only defines these variables and discusses some of their distributional characteristics.

### AGE, SEX, and RACE

Three variables which are excluded from UI profiling are the Age, Sex and Race (Ethnicity) of the claimant. The law specifically prohibits them from being used. Nevertheless, any investigation into the dynamics of exhaustion would be incomplete without some analysis of their impact.

**AGE**—The age of the claimant was calculated using his/her birthdate and the date on which the claimant filed for unemployment. Although the person's nearest age (expressed in years) was calculated, subsequent analysis divided the claimants into two groups—those under 50 years of age and those 50 and over. The mean age of Montana's 1994 claimants was 37.4 years. Sixteen percent were 50 or older.

**SEX**—The sex of the claimant was obtained directly from the BEAR database. The value of 1 was assigned to males, and 0 to females. Of the UI claimants in 1994, 59.5% were men.

**RACE**—An individual's race is also directly available from the BEAR database. Six different categories have been specified. Since Native Americans are Montana's only sizable minority group, three RACE variables were created. WHITE corresponding to BEAR category 1 (White and non-Hispanic), NATIVE corresponding to BEAR category 5 (Native American or Alaskan Native), and OTHER corresponding to the other four BEAR categories.

**Table A1**
**Distribution of Ethnicity among**
**Montana's 1994 UI Claimants**

| ETHNIC GROUP | % TOTAL POPULATION |
|---|---|
| White & Non-Hispanic | 86.7% |
| Native American | 11.2% |
| Other | 2.1% |

### VETERAN

This variable indicates if the claimant claims veteran status. Values are:

1 = Yes
0 = No

About 18.0% of the Montana UI Claimants in 1994 claimed that they were veterans.

## OCCUPATION

An individual's occupation, specified in terms of the nine-digit Dictionary of Occupational Titles (DOT) code, has been an optional field in the BEAR database until June 1995. At that time, it became a mandatory field. Consequently, only about 7% of the three years' records have occupational information. In some cases, there was insufficient data to do even preliminary analyses. See Appendix B.

The first digit of the DOT code was used to obtain the following occupational categories:

1 = Professional, Technical, & Managerial
2 = Sales & Clerical
3 = Services
4 = Agricultural, Fishery, & Forestry
5 = Processing Occupations
6 = Machine Trades
7 = Bench Work
8 = Structural work
9 = Miscellaneous

## PENSION

The amount of a pension the claimant receives, if any, is reported in the BEAR database. Before any analyses were conducted using this variable, it was converted into a categorical variable taking on the following values:

1 = Claimant received some amount from a pension.
0 = Claimant received no amount from a pension.

Less than 0.7% of the claimants received any type of pension.

## SEVERANCE

Whether or not the claimant received any severance pay from his/her last place of employment is information also collected in the BEAR database. Actually, the specific amount received is recorded. In studying this variable, the amount of severance pay was converted into a categorical variable with the following values:

1 = Claimant received some severance pay.
0 = Claimant received no severance pay.

Less than 0.1% of the people filing UI claims in 1994 in Montana received any severance pay.

## MILITARY SERVICE—UCX_EMP

A claimant who has served in the military within the past 18 months is eligible for benefits under the Military's Unemployment Compensation (UCX) program. This item of information was converted into a categorical variable with the following values:

1 = Claimant is eligible for the UCX program.
0 = Claimant is not eligible for the UCX program.

Approximately 3.6% of Montana Claimants in 1994 are ex-military who qualify for UCX reimbursement.

## CIVILIAN EMPLOYEE OF THE FEDERAL GOVERNMENT—UCFE_EMP

A claimant who has been employed as a civilian employee of the federal government in the past 18 months is eligible for benefits under the Federal Government's Unemployment Compensation (UCFE) program. Federal employment was converted into a categorical variable with the following values:

1 = Claimant is eligible for the UCFE program.
0 = Claimant is not eligible for the UCFE program.

In Montana, the Forest Service and Bureau of Land Management use a lot of seasonal employees. Approximately 4.9% of total number of UI Claimants were former federal employees.

## CLAIMANT WORKED IN ANOTHER STATE—CWC_EMP

When a claimant files a Combined Wage Claim (CWC), he/she has worked in more than one state within the past 18 months. These Combined Wage Claims can be expressed as a categorical variable. They take on the following values:

1 = Combined Wage Claim.
0 = Not a Combined Wage Claim.

About 7.4% of the claims filed in Montana were combined wage claims in 1994.

## REPEATED CLAIMANT (MORE THAN ONE SEPARATE CLAIM)—SSN_SEQ

The BEAR database contains a counter which identifies persons who continually make claims year after year. Every new claim is compared to the existing database to see if the claimant has previously established a claim. (Only one claim can be established per benefit year.) If found, the counter is incremented by one. Thus, the counter indicates the number of previous claims (benefit years) established by the claimant.

In UI Profiling, the counter was treated as a categorical variable with the following values:

1 = One or More Previous Claims.
0 = Single Claim.

The 1994 database contains more than 1,800 records (21.0%) for individuals who had filed for unemployment benefits in previous years.

## LABOR SURPLUS AREA—LSA

A Labor Surplus Area is a formal federal designation which makes counties eligible for a variety of governmental programs. To qualify, a county must have an average unemployment rate for the previous two calendar years which is equal to or greater than 120% of the rate for the U.S. and Puerto Rico for the same period. Designations are usually made for the period between Oct. 1 through Sept. 30.

This variable may be considered a substitute for the local unemployment rate (URATE). Because it is categorical and URATE is continuous, one would expect it to have far less predictive power.

Values for claimants living in Labor Surplus Areas are:

1 = County is a Labor Surplus Area.
0 = County is not a Labor Surplus Area.

In 1994, 23.9% of the claimants lived in Labor Surplus Areas.

## COUNTY'S PER CAPITA INCOME—PER_CAP

The per capita income for each claimant's county of residence was obtained from the U.S. Bureau of Economic Analysis. Because of reporting lag, 1992 income was used for the 1994 claimant data. Like unemployment rates it suffers from cluster effects because it is county based and not individual based.

## REPLACEMENT WAGE RATIO—REPLACE_WG

In theory, the Replacement Wage Ratio is the proportion of the claimant's weekly wages which are substituted by the UI benefits. It is calculated by:

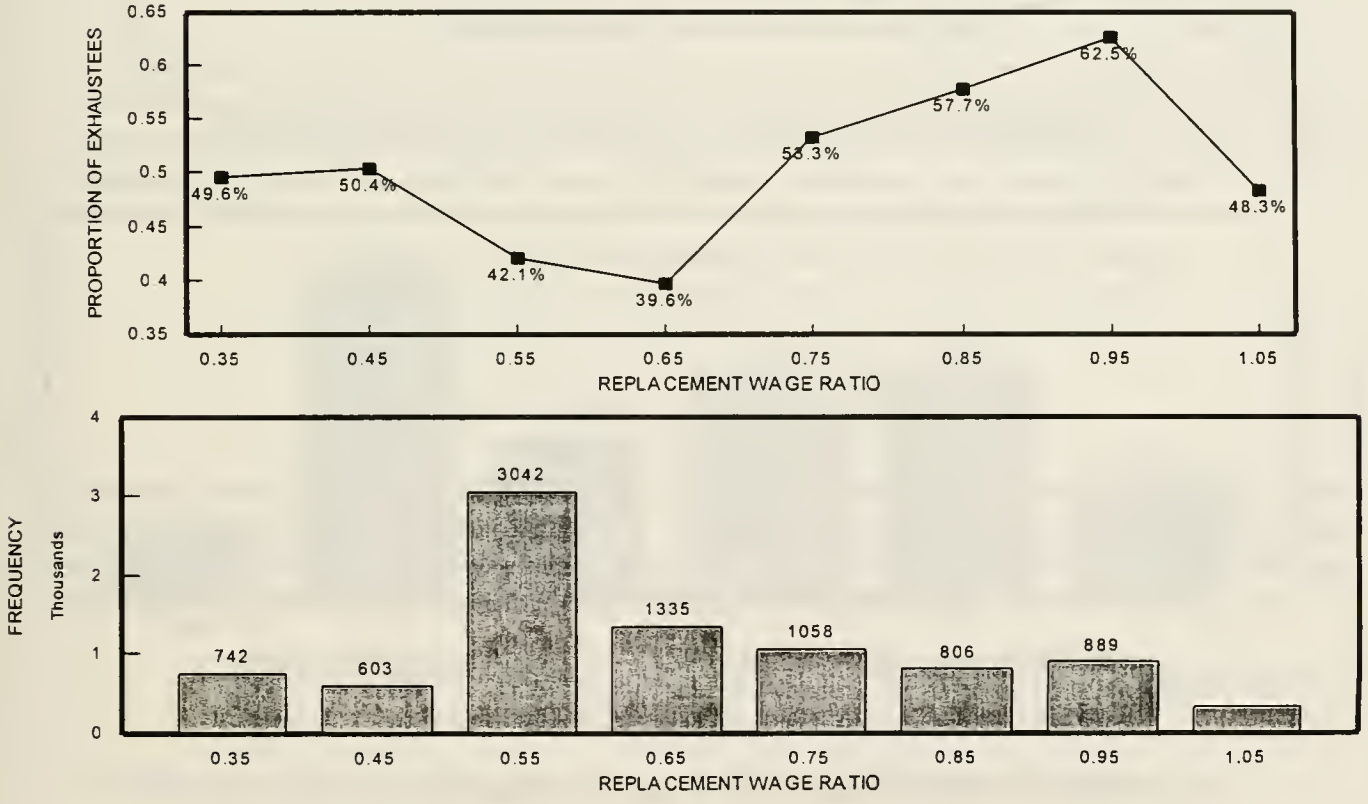
$$\text{Replacement Wage Ratio} = \frac{52 * \text{Weekly Benefit Amount}}{\text{Annual Base Wages}}$$

It is hypothesized that the closer this ratio is to one, the less economic incentive a claimant has to find a job. Therefore, a person who is getting almost as much in weekly UI benefits as take home pay takes longer to find employment and will be more likely to exhaust benefits. The top half of Figure A9 shows that the 1994 data supports this reasoning to some extent.

The above hypothesis assumes that the claimant has worked continuously throughout the base period and that his/her wages were relatively stable. If this were true, this ratio would be around 0.5. (The UI program is designed to replenish a person's wages at this level.) Although we find the most common value for the ratio to be between 0.5 and 0.6, we also find it taking on values ranging from 0.3 to over 2.0. (See the bottom portion of Figure A9.) Such variability is introduced because people are not continuously employed and wage levels do vary from period to period.

Unfortunately, where the replacement wage ratio is the most common, the proportion of people exhausting their benefits is actually dropping. (See the top half of Figure A9.) The proportion of exhaustees does not start to increase until the replacement wage ratio is above 0.6. This phenomenon is precisely what makes this variable a poor predictor.

**Figure A9**
**Proportion of Exhaustees as a Function of the Wage Replacement Ratio**

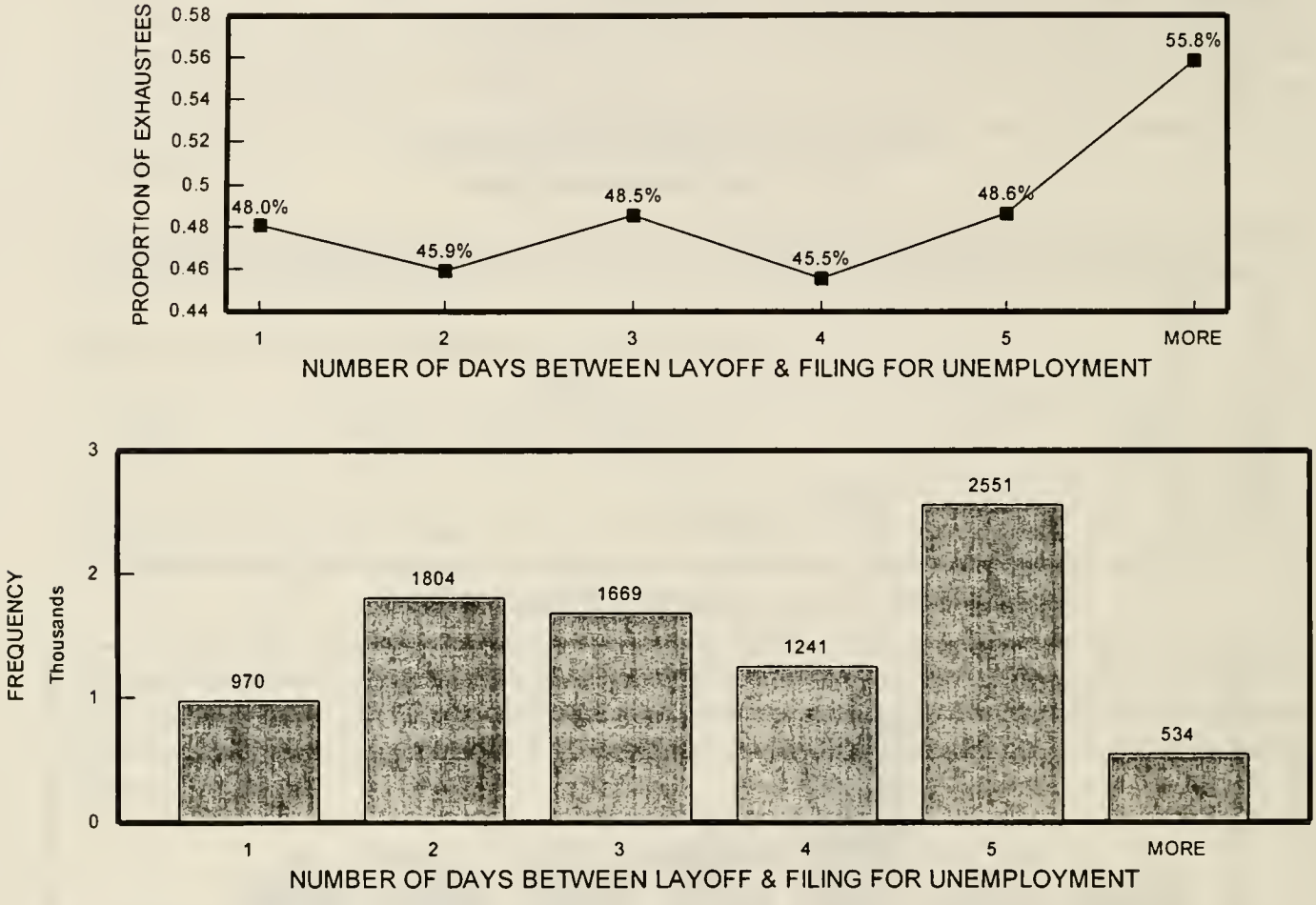



## TIME BETWEEN LAYOFF & FILING—FILE_TIM

FILE_TIM is the delay between the date the claimant lost his/her job and the date when the UI Claim was filed. It is expressed in days. One would think that people delaying the filing of their UI claim will also be less aggressive in their job search, and thus more apt to exhaust their UI benefits.

UI profiling projects in other states have found this variable to be significant. In Montana, however, analysis shows that there is little variation. Only 6% of the people filing UI claims in Montana waited more than 5 days after being laid off before filing. Of the people filing within 5 days, the proportion who exhausted their benefits varied little. See Figure A10.

**Figure A10**
**Proportion of Exhaustees as a Function**
**of the Number of Days taken between Layoff & Filing**




## COUNTIES WITHOUT LOCAL JOB SERVICE OFFICES—LOCAL_OFF

There are 24 job service offices and 56 counties in the state. Yellowstone County is the only county with more than one. Two offices are located in Billings. LOCAL_OFF is a categorical variable constructed for this phenomenon. Having a Job Service Office nearby may be indicative of such factors as the degree of "ruralization," accessibility to employment services, and the magnitude of job opportunities in the area. LOCAL_OFF has the following values:

1 = Local Job Service Office is located in the Claimant's County of Residence.
0 = Local Job Service Office is not located in the Claimant's County of Resi dence.

In 1994, 13.9% of the claimants lived in counties without a local job service office.

## RURAL COUNTIES—RURAL

Rural Counties are defined as counties which have no cities or towns of 2,500 or more population. It is believed that rural areas, having fewer job opportunities, will have higher rates of exhaustion. This variable takes on the following values:

1 = Claimant's County of Residence has at least one town larger than 2,500 people.
0 = Claimant's County of Residence has no town larger than 2,500 people.

Using this definition and 1990 Census data, 30 counties in Montana can be considered rural. Just over 10% of the claimants lived in these 30 counties.

## METROPOLITAN AREAS—METRO

Montana has two metropolitan areas designated by the federal Office of Management and Budget—Yellowstone and Cascade Counties. In contrast to the rural county variable, it is suggested that there are more job opportunities in these areas. The metro variable has the following values:

1 = Claimant's County of Residence is either Yellowstone or Cascade.
0 = Claimant's County of Residence is not Yellowstone nor Cascade.

## SIX MOST POPULOUS COUNTIES—BIG_CTY

Because the definition of a metropolitan area is severely restrictive, an alternative variable to indicate the more populous counties was constructed. BIG_CTY is based upon the six most populated areas of the state: Yellowstone, Missoula, Cascade, Flathead, Gallatin, and Lewis & Clark Counties. Again, BIG_CTY is a categorical variable with the following values:

1 = Claimant's County of Residence is one of the six counties mentioned above.
0 = Claimant's County of Residence is not one of these counties.

In 1994, 56.9% of the claimants lived in one of these "metropolitan" counties.

## WEEKLY BENEFIT AMOUNT—WEEK_AMT

The weekly benefit amount is the dollar amount which the claimant is entitled to receive under the current benefit program. In 1994, the maximum amount was $217 before July 1 and $223 after. In 1994, roughly 23.4% of the claimants were eligible for maximum UI payments.

It is suggested that this variable can be used as a surrogate for income level. This is not completely accurate since Montana employers stop paying UI after the employee has received $15,000 in wages

for the fiscal year. It is hypothesized that people with higher incomes tend to exhaust their benefits more frequently than those having less income. The theory being that it is easier to find a minimum wage or similar job.

Because there are interpretation problems associated with truncated data sets, the weekly benefit amount is defined as a categorical variable. It takes on the following values:

1 = Claimant receives the maximum weekly payment
(in 1994 this would be defined as either $217 or $223).
0 = Claimant receives less than the maximum weekly amount.

## SIZE OF THE LAST EMPLOYER'S BUSINESS (NUMBER OF EMPLOYEES)—CO_EMP

The size of the claimant's last place of employment is reported in terms of the number of employees. This information can be found in the ES 202 database. The number of employees reported in March 1994 is used as a baseline. This number was converted to a categorical variable to separate small companies from large ones. The dividing point was arbitrarily set at five or more employees. Therefore, it is defined as follows:

1 = The claimant's last employer had 5 or more employees.
0 = The claimant's last employer had less than 5 employees.

## TYPE OF INDUSTRY

The type of industry is a series of categorical variables based on the first two digits of the Standard Industry Classification (SIC) Code for the claimant's last employer. These classifications are:

- o Agriculture, Forestry, & Fishing (SIC Codes 01-09)
- o Mining (SIC Codes 10-14)
- o Construction (SIC Codes 15-17)
- o Manufacturing (SIC Codes 20-39)
- o Transportation & Utilities (SIC Codes 41-49)
- o Wholesale Trade (SIC Codes 50 & 51)
- o Retail Trade (SIC Codes 52-59)
- o Finance, Insurance & Real Estate (SIC Codes 60-67)
- o Services (SIC Codes 70-89)
- o Government SIC Codes (91-97)

Using these categorical variables together is another attempt to assess the impact of seasonal workers on exhaustion. Each of these 10 variables took on a value of 1 if the claimant was part of that industry, and 0 otherwise. Since the categories are mutually exclusive, only one of the ten variables had a value of 1 for a given claimant.

## SEASONAL INDUSTRY—SEASONAL

Each two-digit SIC code industry was reviewed for seasonal effects by examining the number of Unemployment Claims filed from month-to-month during the year. An industry was designated as "seasonal" when significant variability was observed. The actual determination was largely judgmental. The values are:

1 = Last Employer was in a "Seasonal" Industry.
0 = Last Employer was not in a "Seasonal" Industry.

Using this largely judgmental variable, 56.2% of the 1994 UI claimants were last employed in a "Seasonal" industry.

## REOPENED CLAIM DURING THE YEAR—SSN_SEQ

SSN_SEQ is the number of times during the benefit year, that the claim has been reopened. This indicates that, for one reason or the other, benefits have not been paid in successive weeks since the benefit year was established. One of the reasons may have been that the claimant found temporary employment. This would indicate an aggressiveness on the part of the job seeker, and one would not expect UI benefits to be exhausted.

For use in the model, SSN_SEQ has been converted into a categorical variable with the following values:

1 = Claimant has filed a UI Claim in at least one previous year.
0 = This is the first time the claimant has filed a UI Claim.

Twenty-two percent of the claimants in 1994 had their UI file closed and then reopened at least once during their benefit year.

## ANNUALIZED BASE WAGES—BASE_WAGES

BASE_WAGES is the total amount of wages paid to the claimant during his/her base period. It is used to determine the claimant's weekly benefit amount (WEEK_AMT). Like WEEK_AMT, BASE_WAGES can also be used as a surrogate for income level. By converting it to a categorical variable based roughly on the median income of wage earners in Montana, one should be able to assess whether or not the people with higher pay scales tend to exhaust their benefits more frequently than those having less income.

In 1994, 43.1% of the claimants had annual wages of less than $10,000.

## IS THE CLAIMANT'S BASE EMPLOYER AND LAST EMPLOYER THE SAME?—SAME_EMP

SAME_EMP is a categorical variable which indicates whether or not the claimant's last employer is the same as the base employer making the largest contributions. In effect, it is a surrogate measure of tenure.

It is constructed by comparing the UI Account Numbers of the two employers. The following values are used:

1 = Last Employer is different from the Base Employer.
0 = Last Employer is the same as the Base Employer.

Two-thirds (66.7%) percent of the 1994 claimants had the same employer at these two times.

## HAS THE CLAIMANT WORKED IN THE SAME INDUSTRY DURING THE PAST YEAR?—SAME_IND

SAME_IND is similar to SAME_EMP but it will contain a few more matches. SAME_IND is a categorical variable which indicates whether or not the claimant was working in the same industry as when his/her base wages were determined. It is constructed by comparing the 4-digit SIC Codes of the two (last employer and the base employer) having paid the largest UI benefits. The following values are used:

1 = Last Industry is the same as the Base Employer's Industry.
0 = Last Industry is different from the Base Employer's Industry.

If the last employer and the base employer were the same (i.e., SAME_EMP = 1), then the value of SAME_IND would also have to be one.

A review of the 1994 records shows that 73.5% of the people filing UI claims worked in the same industry for the past year.

## ARE THE REASONS FOR UNEMPLOYMENT DIFFERENT?—SAME_REAS

SAME_REAS is a categorical variable which indicates whether or not the claimant's reason for being unemployed is the same as the last employer's reason. It is constructed by comparing the two reason codes. The following values are used:

1 = Reasons are different.
0 = Reasons are the same.

Seventy-nine percent of the claimant's explanations given for job loss were the same as the employers' in 1994.

# APPENDIX B
# UNIVARIATE ANALYSIS
# OF
# SELECTED CATEGORICAL VARIABLES

"Univariate" analysis involves each predictor being used alone to test its ability to predict exhaustion. It is usually the first step in statistical modelling. Those variables which are the best individual predictors are usually the ones which have the most predictive value in multivariate models.

An additional benefit to these types of analyses is that one gains some understanding of the factors which influence the phenomenon being studied.

If the predictor variable is categorical, then the effort reduces to contingency analysis. The statistic used is called the odds ratio or "Psi." The odds ratio approximates how much more likely a person has one characteristic over another. For example, if y indicates the presence or absence of cancer and x indicates whether or not a person is a smoker, then an odds ratio of 2 suggests that cancer is twice as likely to occur in smokers than in nonsmokers.

If the odds ratio is 1, there is no predictive value; both outcomes are equally likely. If the odds are less than one, then the absence of the characteristic under study is more likely to occur. Because the software being used automatically generates confidence intervals for Psi, one can identify significant predictors by visual inspection. If the interval contains 1, then it probably is not significant and is probably not a good predictor.

Table B1 summarizes the results of using each of the 33 variables to predict if a claimant in Montana in 1994 would exhaust unemployment insurance benefits. Definitions of these variables can be found in Appendix A.

All variables identified in Table B1 have been converted into categorical variables. For example, instead of using the continuous variable "age," we group the claimants into two categories—"under 50" and "50 and older."

Table B1 presents the predictor variables into four groups. Those describing the claimant's:

- o personal characteristics,
- o his/her job (employment),
- o his/her former employer's characteristics, and
- o descriptors of the geographical area where he/she lives.

With respect to personal characteristics, Age and Race seem to have the largest association with exhaustion. They, however, cannot be used in the Profiling model. The education variables are only mildly sensitive to exhaustion. Veteran's status shows some promise; the very aggregated occupational descriptors do not.

In the group of variables which describe the claimant's employment history, replacement wage, multiple employers, same employer, same industry, and tenure are the best predictors. The latter four, however, all are different measures of job longevity. Two variables, recent military service and weeks of UI eligibility, seem to be negatively correlated to exhaustion. This indicates that people who have recently been released from the military and people who qualify for the full 26 weeks of UI benefits do not exhaust their UI benefits as much as the general population.

The only employer characteristic that appears to be significantly associated with exhaustion is whether or not the claimant has been working in a growth industry. People who work in nongrowth industries have exhaustion rates 33% higher than the people in growth industries.

Seasonality does seem to influence exhaustion, but not in any consistent way. With respect to industry type, only agriculture and retail trade have some influence on exhaustion. People in agriculture have higher exhaustion rates while people in retail trade have lower. On the other hand, seasonality can also be examined from a time perspective. People filing in February and June are less likely to exhaust, while people filing in September and October are more likely.

With respect to geographical indicators, high unemployment rates and being designated a Labor Surplus Area make for higher exhaustion rates. People living in the six most populated areas of the state tend to exhaust at a somewhat lower rate than those in the rest of the state. The other geographical indicators are not significant.

# Table B1
# Odds Ratios (Psi) and 95% Confidence Intervals Showing the Influence of Selected Categorical Variables on Predicting Exhaustion

| VARIABLE | PSI (ODDS RATIO) | 95% CONF INT LOW | 95% CONF INT HIGH |
|---|---|---|---|
| CLAIMANT'S PERSONAL CHARACTERISTICS | | | |
| Age Greater than 50 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX | 1 89 | 1 65 | 2 15 |
| Sex = Male XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX | 1 | 0 91 | 1 09 |
| Race = Non-White XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX | 1 35 | 1 19 | 1 53 |
| Education -- Not a High School Graduate********** | 1 31 | 1 14 | 1 51 |
| -- High School Graduate | 1 1 | 1 01 | 1 2 |
| -- Some College ********** | 0 85 | 0 77 | 0 94 |
| -- College Graduate ********** | 0 87 | 0 77 | 0 97 |
| Non-Veteran ********** | 1 26 | 1 1 | 1 44 |
| Occupation -- Professional Technical & Managerial | 0 8 | 0 67 | 0 96 |
| -- Sales & Clerical | 0 84 | 0 56 | 1 26 |
| -- Services | 0 89 | 0 56 | 1 4 |
| -- Agricultural Fishery & Forestry | 1 16 | 0 48 | 2 8 |
| -- Processing | Insufficent Data | | |
| -- Machine Trade | 2 23 | 1 03 | 4 96 |
| -- Bench Work | 0 15 | 0 01 | 1 26 |
| -- Structural Work | 1 22 | 0 79 | 1 89 |
| -- Miscellaneous | 1 | 0 6 | 1 65 |
| CLAIMANT'S EMPLOYMENT | | | |
| Replacement Wage Ratio > 75 ********** | 1 83 | 1 65 | 2 03 |
| Same Employer (Base vs Last)? ********** | 1 56 | 1 43 | 1 7 |
| Same Industry (Base Emp vs Last Emp ) | 1 55 | 1 42 | 1 7 |
| Three or More Years with the Last Employer | 1 47 | 1 3 | 1 65 |
| Same Reason for Dismissal (Claiment vs Employer) | 1 42 | 1 29 | 1 58 |
| Civilian Employee of the Federal Government within the last 18 Mo***** | 1 39 | 1 14 | 1 7 |
| Worked in no more than one State (Combined Wage Claim)********** | 1 57 | 1 43 | 1 71 |
| Worked for more than 5 employers during base period? ********** | 1 8 | 1 35 | 1 89 |
| Receiving $ from a Pension | 1 26 | 0 78 | 2 08 |
| Weekly Payments = Max Amt (>= $217 & $223) | 1 18 | 1 07 | 1 31 |
| More than One Separate Claim | 1 16 | 1 06 | 1 29 |
| Time Between Layoff & Filing is More than 3 Days ********** | 1 06 | 0 96 | 1 14 |
| Base Wages were <= $9,000 | 0 94 | 0 86 | 1 02 |
| Number of Weeks of UI Eligibility = 26 ********** | 0 78 | 0 72 | 0 85 |
| Military Service within the last 18 Months ********** | 0 65 | 0 52 | 0 83 |
| Received Severance Pay from last Employer | Insufficent Data | | |
| EMPLOYER'S CHARACTERISTICS | | | |
| NonGrowth Industry ********** | 1 32 | 1 21 | 1 44 |
| Last Employer had less than 5 Employees? ********** | 0 96 | 0 89 | 1 12 |
| Type of Industry | | | |
| Agriculture | 1 24 | 1 1 | 1 39 |
| Mining | 1 16 | 0 87 | 1 56 |
| Construction | 1 12 | 0 97 | 1 29 |
| Manufacturing | 1 23 | 1 06 | 1 42 |
| Transportation & Utilities | 0 9 | 0 73 | 1 11 |
| Wholesale Trade | 0 9 | 0 72 | 1 1 |
| Retail Trade | 0 79 | 0 71 | 0 89 |
| Finance, Insurance, & Real Estate | 1 18 | 0 91 | 1 53 |
| Services | 1 | 0 91 | 1 11 |
| Government | 1 23 | 0 95 | 1 59 |
| Seasonal Industry | 1 09 | 0 98 | 1 21 |
| Month in which the Initial Claim was filed | | | |
| January ********** | 1 13 | 1 01 | 1 27 |
| February ********** | 0 7 | 0 59 | 0 83 |
| March ********** | 1 09 | 0 91 | 1 31 |
| April ********** | 0 88 | 0 75 | 1 03 |
| May ********** | 0 88 | 0 75 | 1 03 |
| June ********** | 0 74 | 0 62 | 0 87 |
| July ********** | 0 93 | 0 81 | 1 06 |
| August ********** | 0 97 | 0 79 | 1 19 |
| September ********** | 1 48 | 1 21 | 1 62 |
| October ********** | 1 29 | 1 13 | 1 47 |
| November ********** | 1 11 | 0 94 | 1 3 |
| December | 0 89 | 0 75 | 1 06 |
| GEOGRAPHICAL CHARACTERISTICS (CLAIMANT'S RESIDENCE) | | | |
| Unemployment Rate > 10% ********** | 1 62 | 1 36 | 1 88 |
| Labor Surplus Area | 1 49 | 1 35 | 1 65 |
| Counties without Job Service Offices | 1 16 | 1 02 | 1 31 |
| County's Per Capita Income (1992) < $16,500 | 1 06 | 0 96 | 1 18 |
| Metropolitan Areas (Yellowstone & Cascade Counties) | 0 98 | 0 9 | 1 09 |
| Rural Counties (Counties with no Towns of 2,500 or more) | 0 95 | 0 83 | 1 00 |
| Six Most Populous Counties in MT | 0 8 | 0 73 | 0 87 |

XXXXXXXXXXXXXXXX By Law, these variables cannot be used in the UI Profiling Model

********** Variable was considered for use in the final multivariate model

# APPENDIX C
# THE MULTIVARIATE MODEL

## PRELIMINARY ANALYSES

The multivariate model was developed by using multiple regression and discriminant analysis techniques to identify the most promising predictors of exhaustion. Both used a Step-up approach. Logistic regression could not be used at this initial stage because the amount of memory required was prohibitive.

Twenty-four variables were screened in the initial pass. In rank order, the best predictors were:

- o Weeks of Eligibility (WEEKS_ELG)
- o Years of Employment with the Last Employer (TENURE)
- o Unemployment Rate (URATE)
- o Number of Employers in the Base Period (NUM_EMP)
- o Weekly Benefit Amount (WEEK_AMT)
- o Days between Layoff and Filing (FILE_TIM)
- o Last and Base Period Employers were the Same (SAME_EMP)
- o Month (FEB, APRIL, JULY, and SEPT)
- o Growth Industry (GROWTH)
- o Educational Attainment (ED3)

Review of each of these possible predictors eliminated several. FILE_TIM was dropped for lack of variability; SAME_EMP, because it duplicated TENURE; and WEEK_AMT, because it contributed little to the predictive power of the model when all the other variables were present.

Because four of the months and two of the educational variables were significant, all were added to the final model. They are grouped categorical variables called, "design variables." Design variables require that if one is used, all must be used.

Each of the remaining continuous variables were examined for linearity. In logistic regression linearity in the logit (y in Equation 1-1) is assumed. The correct parametric relationship, or scale, between the predictor and the predicted could be linear, quadratic, nonlinear, or binary. All remaining candidates were found to be linear.

## THE FINAL MODEL

Logistic regression was used to specify the final model. The statistics associated with this model are displayed in Table C1. The Wald Statistic (the ratio of the estimated coefficient to its standard error) is used to determine significance. Since the Wald statistic is normally distributed, a value less than -2.00 or greater than 2.00 would indicate significance.

All four of the continuous variables are significant, as is GROWTH. However, among the educational variables, ED3 (Some College) and ED4 (College Graduate) are significant. Several of the seasonal variables are also significant.

**Table C1**
**Estimated Coefficients, Standard Errors**
**and Wald Statistics for Predicting UI Exhaustees in Montana**

| VARIABLE | Coefficient (Beta) | Standard Error SE(Beta) | Wald Statistic | Probability Beta = 0 |
|---|---|---|---|---|
| Constant | 1.7142 | 0.1467 | 11.6872 | 0.0000 |
| WEEKS_ELG | -0.0788 | 0.0047 | -16.7899 | 0.0000 |
| NUM_EMP | -0.1881 | 0.0174 | -10.7829 | 0.0000 |
| TENURE | 0.0492 | 0.0057 | 8.6323 | 0.0000 |
| URATE | 5.6281 | 0.9546 | 5.8954 | 0.0000 |
| GROWTH | -0.1953 | 0.0457 | -4.276 | 0.0000 |
| ED1 | 0.1213 | 0.076 | 1.5962 | 0.1104 |
| ED3 | -0.1754 | 0.0541 | -3.2453 | 0.0012 |
| ED4 | -0.1461 | 0.0641 | -2.2784 | 0.0227 |
| FEB | -0.3903 | 0.0999 | -3.9052 | 0.0001 |
| MAR | 0.0603 | 0.1047 | 0.576 | 0.5646 |
| APRIL | -0.1617 | 0.0927 | -1.7436 | 0.0812 |
| MAY | -0.1742 | 0.0939 | -1.8556 | 0.0635 |
| JUNE | -0.3251 | 0.0972 | -3.344 | 0.0008 |
| JULY | -0.1169 | 0.0898 | -1.3021 | 0.1929 |
| AUG | 0.0314 | 0.1156 | 0.2711 | 0.7863 |
| SEPT | 0.4018 | 0.1147 | 3.5037 | 0.0005 |
| OCT | 0.1971 | 0.0839 | 2.3486 | 0.0188 |
| NOV | 0.0935 | 0.0964 | 0.9699 | 0.3321 |
| DEC | -0.0915 | 0.0999 | -0.916 | 0.3597 |

## INVESTIGATION OF INTERACTION TERMS

The study of possible interactions was limited to interaction among the continuous variables. The software used to perform the calculations used in logistic regression limited the model to a maximum of 20 predictors. Table C2 shows the results of this effort. The only three interactions which showed any significance at all were:

- o Unemployment and Growth (URATE * GROWTH)
- o Weeks of Eligibility and Number of Employers (WEEKS_ELG * NUM_EMP)
- o Weeks of Eligibility and Tenure (WEEKS_ELG * TENURE)

The test statistic used to measure significance is "G". It is based upon the log-likelihood function and is distributed as Chi-Squared. The relationship between G and the log-likelihood functions is:

G = -2* ((log-likelihood of new model) - (log-likelihood of old model))

All three of the above interactions had high G values and Chi-Square probabilities of less than .05. However, they added little to the prediction of exhaustees when incorporated into the model. Consequently, they are not recommended in the final model.

**Table C2**
**Evaluation of Selected Interactions**
**between Predictor Variables**

| LOGISTIC REGRESSION MODEL | LOG-LIKELIHOOD | G | DEGREES FREEDOM | PROB VALUE |
|---|---|---|---|---|
| Main Effects (WEEKS_ELG, NUM_EMP, TENURE, URATE, & GROWTH | -5825.6089 | | | |
| Add MONTH | -5816.2604 | 18.7 | 11 | 0.0668 |
| Add EDUCATION | -5788.4206 | 74.38 | 3 | 0.0000 |
| | | | | |
| Main Effects (ALL) | -5778.5052 | | | |
| Interactions | | | | |
| WEEKS OF ELIGIBILITY * UNEMPLOYMENT | -5778.5047 | 0.001 | 1 | 0.9748 |
| WEEKS OF ELIGIBILITY * NUMBER OF EMPLOYERS | -5773.2424 | 10.53 | 1 | 0.0012 |
| WEEKS OF ELIGIBILITY * TENURE | -5769.4692 | 18.07 | 1 | 0.0000 |
| WEEKS OF ELIGIBILITY * GROWTH | -5778.0899 | 0.831 | 1 | 0.3621 |
| UNEMPLOYMENT * NUMBER OF EMPLOYERS | -5778.5029 | 0.005 | 1 | 0.9459 |
| UNEMPLOYMENT * TENURE | -5778.1678 | 0.675 | 1 | 0.4114 |
| UNEMPLOYMENT * GROWTH | -5769.8117 | 17.39 | 1 | 0.0000 |
| NUMBER OF EMPLOYERS * TENURE | -5775.4491 | 6.112 | 1 | 0.0134 |
| NUMBER EMPLOYERS * GROWTH | -5777.0134 | 2.984 | 1 | 0.0841 |
| TENURE * GROWTH | -5778.3817 | 0.247 | 1 | 0.6192 |

## VALIDATION OF THE MODEL

Statistically, the model can be validated by dividing the predictions into ten subgroups and comparing the predictions with the observed results. A Chi-square statistic is generated to see if the differences are significantly different from zero. In this case, the recommended model has a Chi-Square of 13.95 with 8 degrees of freedom. (Prob Value = .0830) This means that if there wasn't a fit, this particular arrangement would occur about 8% of the time.

The observed values by decile when using the model on 1992, 1993, and 1994 data are shown in Table C3 and in Figure C1. Although there is no formal statistical test to evaluate the model, a visual inspection shows similar proportions except at the extremes.

In the lowest decile (Number 1 in Figure C3), there is little consistency among the predictions for the three different years. At the same time, only 11 observations for the three years are found here. An examination of these 11 cases indicates that they all have many employers (NUM_EMP) during the base period. All had 9 or more employers. The very small number of observations in this decile group makes yearly comparisons difficult.

In the three deciles where one would expect a preponderance of non-exhaustees (Numbers 2, 3 and 4), the model overestimates the proportion of people exhausting benefits. The proportions of the exhaustees for 1992 and 1993 are much higher than expected.

In the higher deciles (5 through 9), the model consistently predicts those people who are more apt to exhaust benefits. In addition, the probability of exhaustion increases as the decile number increases.

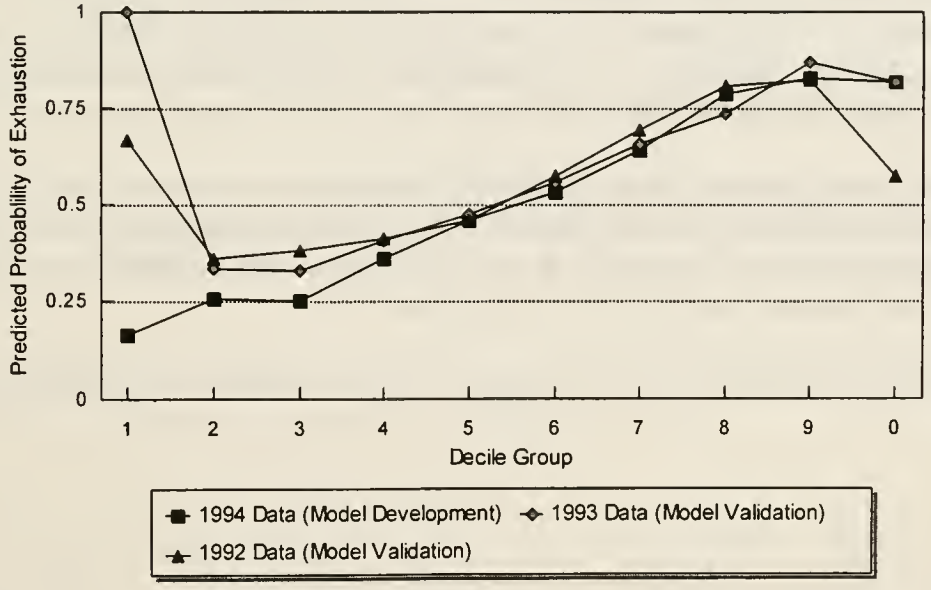
At the top, in decile 10, one would expect 90% or more exhaustees. Unfortunately, there are only 29 observations for the entire three years. Of the 29, we would classify 22 as potential exhaustees. An examination of these records shows that these people are characterized by extremely high values for TENURE.

If we were to evaluate the model in terms of the percentage of exhaustees we would classify correctly, then 60.1% of the 1994 claimants who exhausted their benefits would be classified correctly. Using 1993 data, 61.8% would be classified correctly. For 1992, it is 64.3%. This implies that we should expect at least a 20% improvement over randomly identifying potential exhaustees when using the model.

**Table C3**
**Comparison of the Predicted Number**
**of Exhaustees and Non-Exhaustees**
**when using the UI Profiling Model on Three Different Years of Data**

| | | DECILE | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YEAR | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | TOTAL |
| 1994 | NON-EXHAUST | 5 | 46 | 475 | 1240 | 1359 | 936 | 422 | 82 | 11 | 2 | 4578 |
| | EXHAUST | 1 | 16 | 160 | 706 | 1150 | 1071 | 755 | 304 | 52 | 9 | 4224 |
| | % EXHAUST | 16.7% | 25.8% | 25.2% | 36.3% | 45.8% | 53.4% | 64.1% | 78.8% | 82.5% | 81.8% | 48.0% |
| 1993 | NON-EXHAUST | 0 | 36 | 407 | 1151 | 1396 | 963 | 461 | 130 | 10 | 2 | 4556 |
| | EXHAUST | 2 | 18 | 199 | 797 | 1256 | 1217 | 884 | 355 | 67 | 9 | 4804 |
| | % EXHAUST | 100.0% | 33.3% | 32.8% | 40.9% | 47.4% | 55.8% | 65.7% | 73.2% | 87.0% | 81.8% | 51.3% |
| 1992 | NON-EXHAUST | 1 | 55 | 419 | 1163 | 1396 | 934 | 404 | 89 | 11 | 3 | 4475 |
| | EXHAUST | 2 | 31 | 260 | 816 | 1180 | 1261 | 908 | 366 | 51 | 4 | 4879 |
| | % EXHAUST | 66.7% | 36.0% | 38.3% | 41.2% | 45.8% | 57.4% | 69.2% | 80.4% | 82.3% | 57.1% | 52.2% |

**Figure C1**
**Probability of Exhaustion by Decile Group**



# APPENDIX D
# SENSITIVITY ANALYSIS
# OF THE RECOMMENDED UI PROFILING MODEL
# TO CHANGES IN THE VALUES OF THE PREDICTORS

The sensitivity of the model is defined as the amount of change observed in the predicted values when predictors are changed by a specified, known amount. In this effort, only the marginal change of one predictor at a time will be studied.

To start, a baseline was developed. The following values were used to establish this baseline:

| VARIABLE | VALUE | REASON |
|---|---|---|
| WEEKS_ELG | 26 | 1994 Mode (most common) |
| NUM_EMP | 1 | Mode (most common) which is also positive |
| TENURE | 1 | Mode (most common) which is also positive |
| URATE | 0.063 | 1994 Average |
| GROWTH | 1 | 1994 Mode (most common) |
| EDUCATION | All Zeros | High School Graduate, Mode (most common) |
| SEASON (MONTH) | All Zeros | January is used as a baseline |

Using Equation 1-1 with these values, the probability of exhaustion would be estimated at 48.3%. (The actual proportion of exhaustees in 1994 was 48.0%).

Table D1 shows the results obtained by assuming selected changes in each of the predictor variables. The sign of the coefficients indicates whether the probability would have been increased or decreased.

Reducing the number of weeks a claimant is eligible for UI benefits (WEEKS_ELG) increases the probability of exhaustion by approximately 3.9%. Because of the way in which it is calculated, each unit is reduced by 2. The increase in the probability of exhaustion is relatively stable across the entire range of values. Holding all other independent variables firm, going from 26 weeks to 25 increases the exhaustion probability 3.9%. Going from 14 to 12 weeks, increases it by 3.5%.

Because NUM_EMP (number of employers) and TENURE (years of service with last employer) are whole numbers with limited variability, the model examined the output with values changing from 1 to 2. With respect to tenure, this results in a 1.2% increase in the probability of exhaustion. With NUM_EMP, the increase resulted in a 4.5% decrease.

Changing GROWTH to NONGROWTH results in a 5.8% increase in the probability of exhaustion.

The educational and seasonal variables were changed one at a time. If the claimant didn't have a high school education, his/her probability of exhaustion would be 2.9% higher than one with a high school diploma. Some college results in more of a decrease than a college degree (4.2% vs. 3.5%).

The baseline assumed that the claim was filed in January. If, instead, it was filed in February, then the probability of exhaustion would be 9.1% lower. If it were in September, it would be 10.0% higher.

**Table D1**
**Changes in the Probability of Exhaustion**
**Resulting from Assumed Changes in the Predictor Variable**

| VARIABLE | BASELINE VALUE | FINAL VALUE | NET CHANGE (%) | NET RESULT IN PROB (EXHAUSTION) |
|---|---|---|---|---|
| WEEKS_ELG | 26 | 24 | -7.7% | +3.9% |
| NUM_EMP | 1 | 2 | +50.0% | -4.5% |
| TENURE | 1 | 2 | +50.0% | +1.2% |
| URATE | 0.063 | 0.07 | +10.0% | +1.0% |
| GROWTH | 1 | 0 | | +5.8% |
| ED1 | 0 | 1 | | +2.9% |
| ED3 | 0 | 1 | | -4.2% |
| ED4 | 0 | 1 | | -3.5% |
| FEB | 0 | 1 | | -9.1% |
| MAR | 0 | 1 | | +1.5% |
| APRIL | 0 | 1 | | -3.9% |
| MAY | 0 | 1 | | -4.2% |
| JUNE | 0 | 1 | | -7.7% |
| JULY | 0 | 1 | | -2.8% |
| AUG | 0 | 1 | | +0.8% |
| SEPT | 0 | 1 | | +10.0% |
| OCT | 0 | 1 | | +4.9% |
| NOV | 0 | 1 | | +2.2% |
| DEC | 0 | 1 | | -2.2% |

# REFERENCES


Dallal, G. E. ***LOGISTIC: A Logistic Regression Program for the IBM PC,*** The American Statistician, 42, 272. 1988.

Dean, Andrew G. et al. ***Epi Info, Version 6. A word-processing, database, and statistics program for public health on IBM-compatible microcomputers.*** The Division of Surveillance and Epidemiology, Epidemiology Program Office, Center for Disease Control and Prevention (CDC), Atlanta, GA 30333, April 1994.

Hintze, Jerry L. ***Number Cruncher Statistical System. Version 5.03 Installation and Reference Manual.*** Kaysville, UT 84037, September 1992.

Hosmer, David W. and Stanley Lemeshow. ***Applied Logistic Regression.*** John Wiley & Sons, New York, NY, 1989.

Kennedy, Daniel W. ***The Connecticut UI Profiling Model: A Nonmathematical Presentation.*** Unpublished Working Paper. February 1995.

Kennedy, Daniel W. ***The Connecticut UI Profiling Model: Executive Summary.*** Unpublished Working Paper. March 1995.

Schlauch, Kurt and Marisa Puglisi. ***Worker Profiling and Reemployment Services. Profiling Methods: Lessons Learned.*** Unpublished Working Paper. Undated.

U.S. Department of Labor, Employment and Training Administration. ***The Worker Profiling and Reemployment Services System: Legislation, Implementation Process and Research Findings.*** Unemployment Insurance Occasional Paper Series #94-4. 1994.

150 copies of this public document were published at an estimated cost of $1.48 per copy, for a total cost of $222.00, which includes $222.00 for printing and $.00 for distribution.